International Journal of Computer Science
and Engineering (IJCSE)
ISSN(P): 2278-9960; ISSN(E): 2278-9979
Vol. 3, Issue 2, Mar 2014, 141-150

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

ANALYSIS OF CUSTOMER SATISFACTION FOR COMPETITIVE ADVANTAGE USING CLUSTERING AND ASSOCIATION RULES

IBUKUN AFOLABI & FISAYO ADEGOKE

Department of Computer and Information Sciences, Covenant University, Ota, Nigeria

ABSTRACT

Customer satisfaction is a very important factor in organizational profit and positioning for effective competitive advantage requires making decisions based on quality inferences from data mining. The aim of this paper is to provide competitive advantage inferences based on analyzing customer satisfaction data using the combination of k-means clustering and association rule mining technique. Based on the information gotten from the questionnaires administered to retrieve customer satisfaction information of mobile network service providers in Nigeria, prediction is done and inferences are generated with the help of clusters and association rules. This paper proposes an effective method to extract knowledge from questionnaire data which is very useful for improving the competitive advantage of organizations. In conclusion, the paper has been able to identify the factors that contribute to customer satisfaction in the Nigeria Mobile Network sector.

KEYWORDS: Competitive Advantage, K-Means Clustering, Association Rule Mining, Data Mining Customer Satisfaction

1. INTRODUCTION

Customer satisfaction is the outcome felt by those that have experienced a company's performance that have fulfilled their expectations. Research has revealed that satisfaction has a positive effect on organization's profitability (Angelova & Zekiri, 2011).

Competitive intelligence (CI) on the other hand, involves analyzing the industry in which a firm operates as inputs to the firms strategic positioning, marketing activities and understanding competitor vulnerabilities for better decision making. (Strauss et al. 2006). Competitive intelligence (CI) is a specialized branch of Business Intelligence. It involves a systematic and ethical program for gathering, analyzing and managing external information that can affect the organization's plans, decisions and operations. Customer Relationship Management (CRM) targets markets (customers) while competitive intelligence targets markets (customers) through industrial opportunities. Currently, the stage of development in competitive intelligence can be characterized as "Competitive Intelligence for Strategic Decision Making." Also the future rests on developing CI as a source of competitive advantage. Therefore, a company has competitive advantagewhenever it has an edge over rivals in securing customers and defending against competitive forces.

The telecommunications industry in Nigeria, when established were supposed to provide the following services to mention a few, provide and operate public payphone, provide and operate private network links, employing cable, radio communications or satellite, within Nigeria (Nnama, 1999). According to Roger (2010), there are five GSM network and 13 CDMA-based network operators in Nigeria. The GSM operators include Airtel, MTN, Globacom and Etisalat while the CDMA network operators include Multilinks, Starcoms, Visafone amongst others. Study reveals promotion as a major

marketing management tool for survival, sustenance and expansion of business in the Nigerian Telecommunication in the Nigerian telecommunication industry in Nigeria (Obasan & Soyebo, 2012). Also in Nigeria studies have explored the relationship between strategic agility and competitive advantage in Nigeria's telecommunication industry, revealing that strategic agility influences the competitive performance of telecommunication firms in Nigeria (Oyedijo, 2012).

This paper therefore aims at analyzing Customer Satisfaction data for Competitive Advantage using data mining techniques; k-means clustering and Association Rule Mining.

2. LITERATURE REVIEW

Data mining has changed the sales target of CRM systems from products to customers: How to classify customers? How to find out the common character of customers from database? How to dig up the potential customers? How to find out the most valuable customers? These kinds of questions become the most popular data mining applications in marketing (Xiaoshan, 2006). These data mining techniques include clustering(k-means), Association rule mining, Neural network and many more.

Presently, regarding the application of k-means clustering for purpose of competitive intelligence, the following are the researches that have been carried out to mention a few; (Satish et al., 2012), used k-means clustering for B2B Segmentation using Customers' Perceptions. In their work three clustering algorithms, were compared; K-means, Normal Mixtures and Probabilistic-D. It was discovered that K-means follow a deterministic approach in calculating cluster membership; clustering techniques like Normal Mixtures calculate a degree of membership or probability for each customer to belong to a cluster while the probabilistic-D technique calculates probability of cluster membership using the Euclidean distance of each observation from cluster centres found by k-means. The result showed that there can be a better understanding of markets by using soft clustering techniques.

The analysis of customer service choices and promotion preferences using k-means algorithm was carried out in (Charles, 2009). The study was able to demonstrate that complex menu selections in franchise restaurant can be better managed and promoted. In (Wang and Zhang, 2004), K-means was used for business intelligence purposes. In their work, they were able to propose a KBSVM (KMeans-based Support Vector Machine) method and reveal through experiments that the KBSVM method can build much more succinct model without any significant degradation of the classification accuracy. And finally, K-means clustering method is used to discover knowledge that come from CRM data in a study that identified insurance products and improved product selling strategies (Balajiand Srivatsa, 2012).

Association rule mining on the other hand has been used in various ways to achieve business intelligence; mining changes in patent trends for competitive Intelligence is one of such for example(Shih, 2008), in this research, the change mining approach used, generated competitive intelligence, which is used to help managers develop appropriate business strategies. (Karanikas, 2002) applied temporal text mining in Competitive Intelligence for the biotechnology and pharmaceutical industry, in order to identify changes and trends of associations among entities of interest that appear in text over time. According to (Mert et. al., 2011), competitive advantage can be created by using association rule mining techniquefor decision making. (Shaw et al., 2001) gives a detailed procedure of how the framework of marketing knowledge management can benefit from association rule mining.

Finally, the combination of k-means clustering and Association rule mining approach which is being proposed to be used in analyzing the data has also been used rarely. (Isakki & Rajagopalan, 2012) proposed an effective method to

extract knowledge from transactions records which is very useful for increasing the sales. Customer details are segmented using k-means and then Apriori algorithm is applied to identify customer behaviour.

3. METHODOLOGY

Data mining is the process of discovering unknown patterns in databases. It involves using one or more algorithms, including neural network algorithms, tree induction algorithms, and/or clustering algorithms, association rule mining just to mention a few to identify hidden patterns in the data (Lauría & Peter, 2004). In this paper, the customer satisfaction data with similar attribute are first grouped by means of clustering techniques. Finally, for each cluster, an association rules are used to identify the products that are frequently bought together by the customers.

3.1 K-Means Algorithm

According to (Satish et al., 2012), K-means algorithm is one of the most widely used hard clustering techniques.

Clustering algorithms can be partitional or hierarchical. The k-means clustering is such that requires no prior knowledge of relationships. In this paper, we apply the k -means algorithm to segment questionnaires used to retrieve customer satisfaction information regarding Nigerian mobile network service providers.

The algorithm works as follows:

- Specify the number of clusters (k in k-means)
- Randomly select k cluster centres in the data space
- Assign data points to clusters based on the shortest Euclidean distance to the cluster centres
- Re-compute new cluster canters by averaging the observations assigned to a cluster.
- Repeat above two steps until convergence criterion is satisfied.

The advantage of k-means clustering is that it can handle large data sets and can work with compact clusters (Satish et al., 2012).

3.2 Association Rule Algorithm

Given a set of keywords $\{\ \}\ A = w1\ ,w2\ ,...,\ wn$and a collection of indexed documents $D = \{d1,d2,...,\ dm\}$, where each document $i\ d$ is a set of keywords such that $di \subseteq A$. Let Wibe a set of keywords. A document $i\ d$ is said to contain Wi if and only if $Wi \subseteq di$. An association rule is an implication of the form $Wi \Rightarrow Wj$ where $Wi \subset A$, $Wj \subset A$ and $Wi \cap Wj = \varphi$ (Hany, 2007).

Association rule mining algorithm makes use of two measures; support(s) and confidence(c). The rule $Wi \Rightarrow Wj$ has support s in the collection of documents D if $s\%$ of documents in D contain $Wi \cup Wj$. The support is calculated by the following formula:

$$Support(W_i W_j) = \frac{SupportCount(W_i W_j)}{TotalNumberOfTransac\ tions}$$

The rule $Wi \Rightarrow Wj$ holds in the collection of documents D with confidence c if among those documents that contain Wi, c % of them contain Wj also. The confidence is calculated by the following formula: (Hany, 2007)

$$Confidence(W_i W_j) = \frac{Support(W_i W_j)}{Support(W_i)}$$

The algorithm for generating association rules based on the weighting scheme is given as follows (Hany, 2007):

- Scan the file that contains all the keywords that satisfy the threshold weight value and their frequency in each document.
- Let N denote the number of top keywords that satisfy the threshold weight value.
- Store the top N keywords in index file along with their frequencies in all documents, their weight values relevance Weight and documents ID in the following format: <doc-id><keyword><keyword frequency><relevanceWeight>
- Scan the indexed file and find all keywords that satisfy the threshold minimum support. These keywords are called large frequency1-keyword Set L_1.
- When K is greater than 2, (Note K is a keyword set having k-keywords sets). The candidate keywords C_k of size K are generated from large frequent (k-1) keywords sets, L_{k-1} that is generated in the last step.
- Scan the index file, and compute the frequency of candidate keyword sets C_k that is generated in step 4.
- Compare the frequencies of candidate keywords sets with minimum support.
- Large frequent keyword sets L_k, which satisfy the minimum in support, is found from step 7 above.
- For each frequent keyword set, find all the association that satisfies the threshold minimum confidence.

3.3 Data Description

For this research in competitive intelligence, data was gathered using questionnaire, designed and administered to 200 respondents. The questionnaire contained both structured and unstructured part. The structured part of the questionnaire consists of demographic profile of respondents such as gender, age, academic qualification, occupation, state and nationality. This is important to have background information about the respondents.

Other items in the structured part of the questionnaire have to do were the mobile phone and network usage of the respondents. Respondents were asked to respond to questions on their mobile phone usage such as voice calls, data and SMS on a five point like scale ranging from strongly agree to strongly disagree. Participants were asked to rate performance the customer service of their network whether good, satisfactory, unsatisfactory or poor.

The unstructured part includes the respondents' answers to questions such as:

- What do you like most about your network service?
- What do you dislike most about your network service?
- What improvements would you like to see, if any with regard to your network service?
- What type of problem do you usually encounter with your network service?

3.5 Weka Work Bench

Weka (Waikato Environment for Knowledge Analysis) is a Java-based data mining tool developed by Waikato University. I loads data, preprocesses the data. This preprocessing includes information extraction stages like stemming ndstopword removal in the case of unstructured data). In the filtration process, the unstructured data(text) are filtered by removing the unimportant words from documents content. Such unimportant words include: articles, pronouns, determiners, prepositions and conjunctions, common adverbs and non-informative verbs. After the filtration process, the extracted words are stemmed which is the process that removes a word's prefixes and suffixes (such as unifying both infection and infections to infection). The algorithms available in Weka includes; classification, Clustering, and Association (Hall et al., 2008).

4. RESULTS OF THE STUDY

The result of this study is presented in three different parts, the first part reports the output of the simple k-means algorithm, the second part is the output of the association rule mining while the last section is the combination of the k-means and association rule mining output in order to make competitive advantage recommendations.

4.1 K-Means

Simple k-means algorithm is then applied. The following information are the results of Weka tool.

=== Run information ===

Scheme: weka.clusterers. Simple K Means -N 3 -A "weka. core. Euclidean Distance -R first-last" -I 50 -num-slots 1 -S 30

Relation: ClusterData2-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate2.0-N0-S-stemmerweka.core.stemmers.NullStemmer-M3-O-tokenizerweka.core.tokenizers.WordTokenizer-delimiters"

\r\n\t.,;:\'\"()?!"

Instances: 200

Attributes: 98

Test Mode: evaluate on training data

Number of Iterations: 8

Within Cluster Sum of Squared Errors: 1761.1800471559097

Missing values globally replaced with mean/mode

Time Taken to Build Model (Full Training Data): 0.25 seconds

Clustered Instances

c0 58 (29%)

c1 52 (26%)

c2 90 (45%)

Table 1: Selected Attributes

c0	c1	c2
sms	sms	female
voicecalls	voicecalls	sms
data	male	voicecalls
male	data	data
lagos	poor	mtn
5000	network	lagos
bsc	mtn	21-30
student	31-40	student
satisfactory	lagos	bsc
21-30	satisfactory	poor
15-20	5000	blackberry
mtn	call	network
poor	nokia	5000
service	msc	satisfactory
blackberry	service	1000
internet	ogun	coverage
customer	lecturer	call
services	10000	cheap
nokia	5y	wide
10000	rates	service

Table 1 reveals the 20 items having the highest weight values in the three clusters viewed in Microsoft excel after sorting the attributes according to the weight of their occurrence in the clusters.

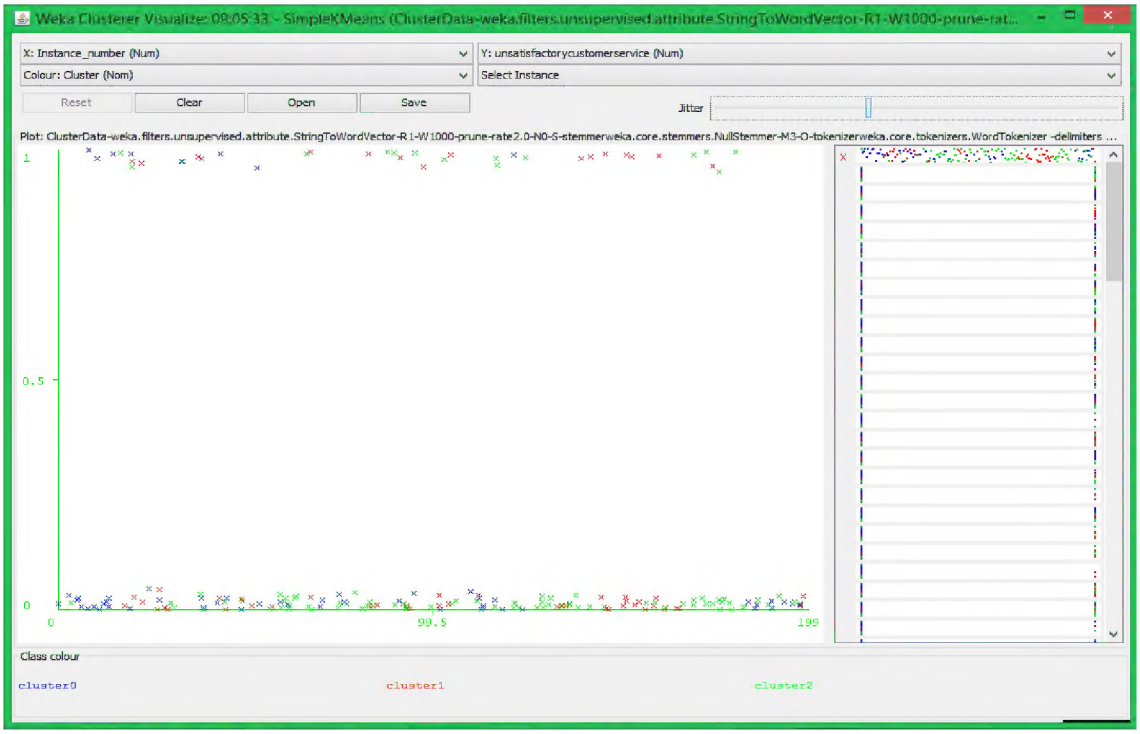

Figure 1: Visualization of the Clusters

The clusters presented above is visualized in Figure 1, it reveals three clusters such that cluster0 is blue colour, cluster1 is red and cluster2 is green. The figure is a graph of the instance number verse a particular variable. In this case, the instance is *unsatisfactorycustomerservice* and its revealed to appear in all the in clusters with an instance number of above 0.9.

4.2 Association Rule

Apriori algorithm is then applied. The following informations are the result of Weka tool.

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 0.9 -M 0.09 -S -1.0 -c -1

Relation: associationdata-weka.filters.unsupervised.attribute.StringToNominal-Rfirst-last

Instances: 200

Attributes: 26

=== Associator model (full training set) ===

Minimum Support: 0.09 (18 instances)

Minimum metric <confidence>: 0.9

Number of Cycles Performed: 19

Generated Sets of Large Itemsets:

Size of Set of Large Itemsetsl(1): 38

Size of Set of Large Itemsetsl(2): 21

Size of Set of Large Itemsetsl(3): 3

Table 2: Best Rules Found

Best Rules
col2=wide 34 ==> col3=coverage 34 <conf:(1)>
col13=voicecalls col14=data 24 ==> col15=sms 24 <conf:(1)>
col1=mtn col3=coverage 22 ==> col2=wide 22 <conf:(1)>
col1=mtn col2=wide 22 ==> col3=coverage 22 <conf:(1)>
col12=voicecalls col14=sms 18 ==> col13=data 18 <conf:(1)>
col12=voicecalls col13=data 18 ==> col14=sms 18 <conf:(1)>
col3=coverage 35 ==> col2=wide 34 <conf:(0.97)>
col14=data col15=sms 25 ==> col13=voicecalls 24 <conf:(0.96)>
col13=voicecalls col15=sms 25 ==> col14=data 24 <conf:(0.96)>
col14=data 27 ==> col15=sms 25 <conf:(0.93)>

4.3 Combining K-Means and Association Rule Mining

Comparing Table 1 and Table 2, we discovered that the rule *col2=wide 34 ==> col3=coverage 34* is one of the best rules, and this rule is clearly the focus of cluster c2 which is the biggest cluster with 90 instances making 45% of the total instances. The rule therefore gives more understanding and interpretation to cluster c2. Making inference from the combination of this rule and this cluster clearly indicates that *female students*, who uses their phone for *sms*, *voicecalls* and

data, who live in *lagos*, have at least *B.Sc* degree, are between the ages of *20 and 30* years old use *mtn* (as their mobile network service provider) particularly with a *blackberry* mobile device. This set of customers experience *poorservice* (network) though they are *satisfied* with the *customer service* of *mtn*. Also this category of users believes that *mtn* has wide *service coverage*. Using this information for competitive advantage means that operators of other mobile network service such as *glo, airtel, visaphone, etisalatetc* (which never feature in the Table 1 and Table due to low frequency) can target these sector of customers (*female students*) and provide *wide* network *coverage* to be able to increase their customer base thereby improving their overall profit.

5. CONCLUSIONS

In conclusion, customer satisfaction is a major part of contributing to the profit of the organization and this can be achieved through competitive intelligence. In this research we have used the combination of Association rule mining and k-means clustering to make competitive advantage based inferences.

Findings from the system reveal that there is strong relationship between a particular sector of customers, which are the female students and some attributes of customer satisfaction in this sector which includes, network coverage and customer service.

Finally, using association rules in combination with k-means as opposed to the traditional statistical analysis has helped to reveal unique interesting relationships among items in the data received from the questionnaire.

6. REFERENCES

1. Angelova B. and Zekiri J. (2011) "Measuring Customer Satisfaction with Service Quality Using American Customer Satisfaction Model (ACSI Model)" *International Journal of Academic Research in Business and Social Sciences, Vol. 1, No. 3 ISSN: 2222-6990*
2. Balaji S. and Srivatsa S.K. (2012) "Data Mining model for Insurance Trade in CRM System" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, April 2012 ISSN: 2277 128X Available online at: www.ijarcsse.com
3. Charles V. T.(2009)The Analysis Of Customer Service Choices andPromotion Preferences using HierarchicalClustering, *Journal of the Chinese Institute of Industrial Engineers, Vol. 26, No. 5, pp. 367-376*
4. Hall M. Eibe Frank, Holmes G., Pfahringer B. (2008). The WEKA Data Mining Software: An Update in SIGKDD Explorations Volume 11, Issue 1 Page 10-13
5. Hany M.; Dietmar R.; Nabil I. and Fawzy T.(2007) "A Text Mining Technique Using Association Rules Extraction", International Journal of Computational Intelligence volume 4 number 1 2007 ISSN 1304-2386.
6. Isakki P., and Rajagopalan S.P.(2012)Analysis of Customer Behavior using Clustering and Association Rules, *International Journal of Computer Applications (0975 – 8887) Volume 43– No.23*
7. Karanikas H.., Koundourakis G., Kopanakis I. and Mavroudakis T. (2002). A Temporal Text Mining Application in Competitive Intelligence.

8. Lauría E.J. and Peter J. D., (2004) "A Bayesian Belief Network-driven Decision Support System for Client-Server Implementation" Decision Support in an Uncertain and Complex World: The IFIP TC8/WG8.3 International Conference pp 447-456.
9. Mert B., Yasemin B. and Ayşe D. (2011) "Creating competitive advantage by using data mining technique as an innovative method for decision making process in business" in *Annual Conference on Innovations in Business & Management London, UK*
10. Nnama E. C. (1999) "Present Status of the Telecommunications Industry in Nigeria – A Brief on the Nigerian Communications Commission (NCC)".

    http://www.ncc.gov.ng/archive/Archive%20of%20Speeches/Present%20Status%20of%20Telecoms.pdf
11. Obasan K. A. and Soyebo Y. A. (2012) "Assessing the Effectiveness of Promotion as a Marketing Management Tool in the Nigerian Telecommunication Industry" Journal of Emerging Trends in Economics and Management Sciences (JETEMS) 3 (1): 1-6
12. Oyedijo A. (2012)Strategic Agility and Competitive Performance in the Nigerian Telecommunication Industry: an Empirical Investigation.Business and Management Review Vol. 1(12) pp. 39 – 50 ISSN: 2047 - 0398 Available online at http://www.businessjournalz.org/bmr.
13. Roger, F. (2010). *Nigeria rises* (Internet).http://www.arabianbusiness.com/580722nigeriarises 9th February 2010.
14. Satish G., Goutam C., and Gary G.(2012) Comparison of K-means, Normal Mixtures and Probabilistic-D Clustering for B2B Segmentation using Customers' Perceptions in SAS Global Forum.
15. Shaw M. J., Subramaniam C, Tan G. W. and M. E. Welge, "Knowledge management and data mining for marketing", Decision Support Systems, v.31 n.1, pages 127-137, 2001.
16. Shih, Meng-Jung; Liu, Duen-Ren; Hsu, Ming-Li (2008): Mining Changes in Patent Trends for Competitive Intelligence.
17. Strauss, J., El-Ansary, A., and Frost, R., "E-marketing", International Edition, Published by Pearson Prentice Hall, pp30, 135-168, 2006.
18. Wang J. and Zhang C.(2004)KBSVM: KMeans-based SVM forBusiness Intelligence in Americas Conference on Information Systems (AMCIS)AMCIS 2004 Proceedings
19. Xiaoshan D., "Data Mining Analysis and Modeling for Marketing Based on Attributes of Customer Relationship", Reports from MSI- Rapporterfrån MSI, Report 06129 Växjö University ISSN 1650-2647, 2006. pp 1-46 www.msi.vxu.se